NeuroFL

A Federated Data Network for Brain Health

TPC26 FL - June 2026 – Francis Jeanson, PhD

ONTARIO BRAIN INSTITUTE
INSTITUT ONTARIEN DU CERVEAU

Ontario Brain Institute (OBI)

WHO WE ARE

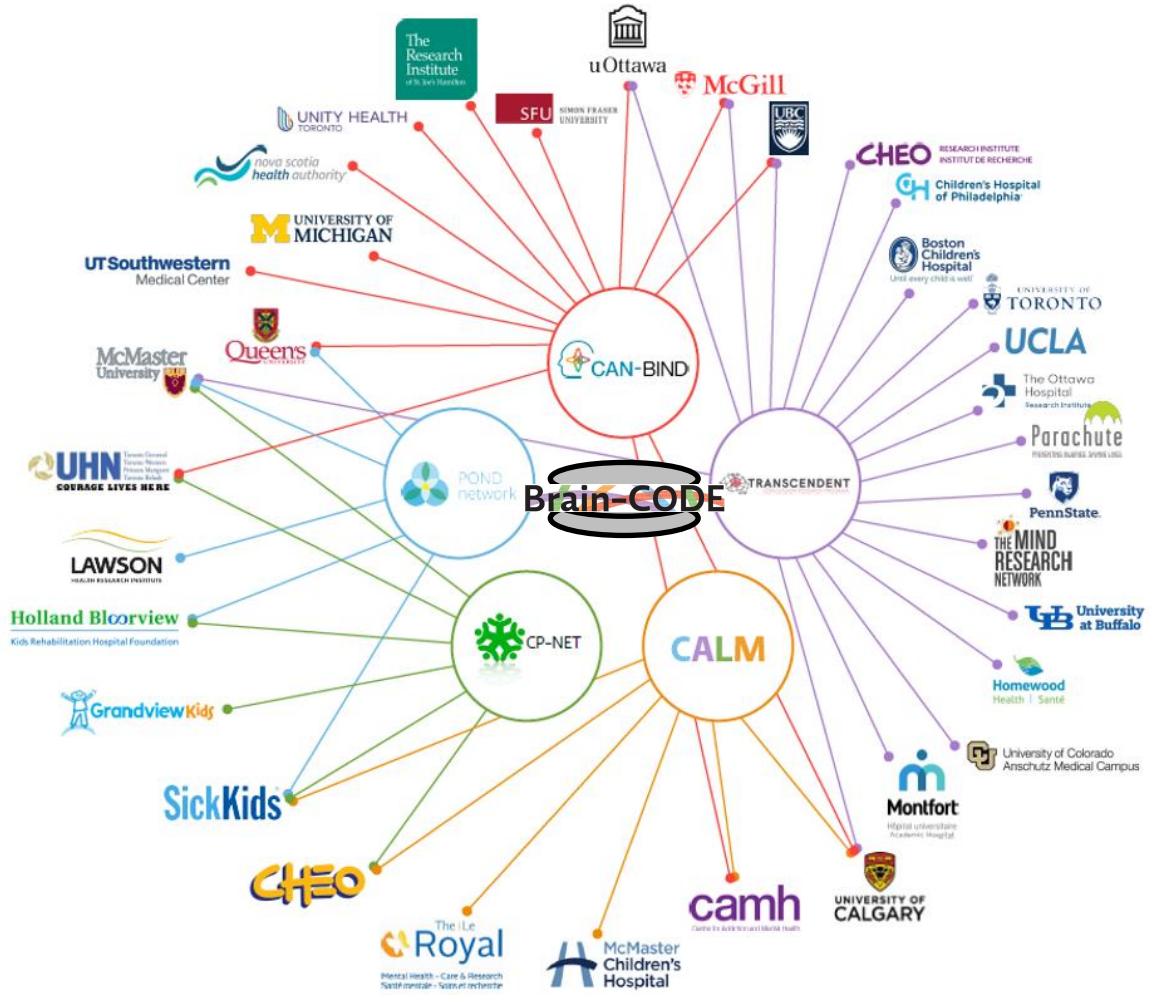
A provincially and federally funded, not-for-profit that accelerates discovery and innovation, benefitting both patients and the economy

VISION

Make Ontario a world leader in brain research, commercialization and care

CORE PURPOSE

Accelerate solutions to improve brain health, through a collaborative approach to research, commercialization, and knowledge translation



Brain-CODE

- 30,000+ participants
- 40+ Institutions
- 1000+ Brain-CODE Users
- Data Standardization
- Data Federation
- Data Linking
- Data Sharing (FAIR)

Ontario Brain Institute
Centre for Analytics

Predicting Drug Response for Major Depressive Disorder

Psychological Medicine

cambridge.org/psm

Original Article

Cite this article: Sajjadian M et al (2022). Prediction of depression treatment outcome from multimodal data: a CAN-BIND-1 report. *Psychological Medicine* 1–11. https://doi.org/10.1017/S0033291722002124

Received: 10 December 2021
Revised: 4 May 2022
Accepted: 20 June 2022

Key words:
Machine learning; MDD; predictive analysis; treatment outcome

Author for correspondence:
Rudolf Uher,
E-mail: uher@dal.ca

Prediction of depression treatment outcome from multimodal data: a CAN-BIND-1 report

Mehri Sajjadian[1], Rudolf Uher[1], Keith Ho[2,3], Stefanie Hassel[4,5], Roumen Milev[6], Benicio N. Frey[7,8], Faranak Farzan[9], Pierre Blier[10,11,12], Jane A. Foster[13], Sagar V. Parikh[14], Daniel J. Müller[15,16], Susan Rotzinger[16,17], Claudio N. Soares[18], Gustavo Turecki[19], Valerie H. Taylor[20], Raymond W. Lam[21], Stephen C. Strother[22,23] and Sidney H. Kennedy[16,17,24,25]

Abstract

Background. Prediction of treatment outcomes is a key step in improving the treatment of major depressive disorder (MDD). The Canadian Biomarker Integration Network in Depression (CAN-BIND) aims to predict antidepressant treatment outcomes through analyses of clinical assessment, neuroimaging, and blood biomarkers.

Methods. In the CAN-BIND-1 dataset of 192 adults with MDD and outcomes of treatment with escitalopram, we applied machine learning models in a nested cross-validation framework. Across 210 analyses, we examined combinations of predictive variables from three modalities, measured at baseline and after 2 weeks of treatment, and five machine learning methods with and without feature selection. To optimize the predictors-to-observations ratio, we followed a tiered approach with 134 and 1152 variables in tier 1 and tier 2 respectively.

Results. A combination of baseline tier 1 clinical, neuroimaging, and molecular variables predicted response with a mean balanced accuracy of 0.57 (best model mean 0.62) compared to 0.54 (best model mean 0.61) in single modality models. Adding week 2 predictors improved the prediction of response to a mean balanced accuracy of 0.59 (best model mean 0.66). Adding tier 2 features did not improve prediction.

Conclusions. A combination of clinical, neuroimaging, and molecular data improves the prediction of treatment outcomes over single modality measurement. The addition of measurements from the early stages of treatment adds precision. Present results are limited by lack of external validation. To achieve clinically meaningful prediction, the multimodal measurement should be scaled up to larger samples and the robustness of prediction tested in an external validation dataset.

CAN-BIND

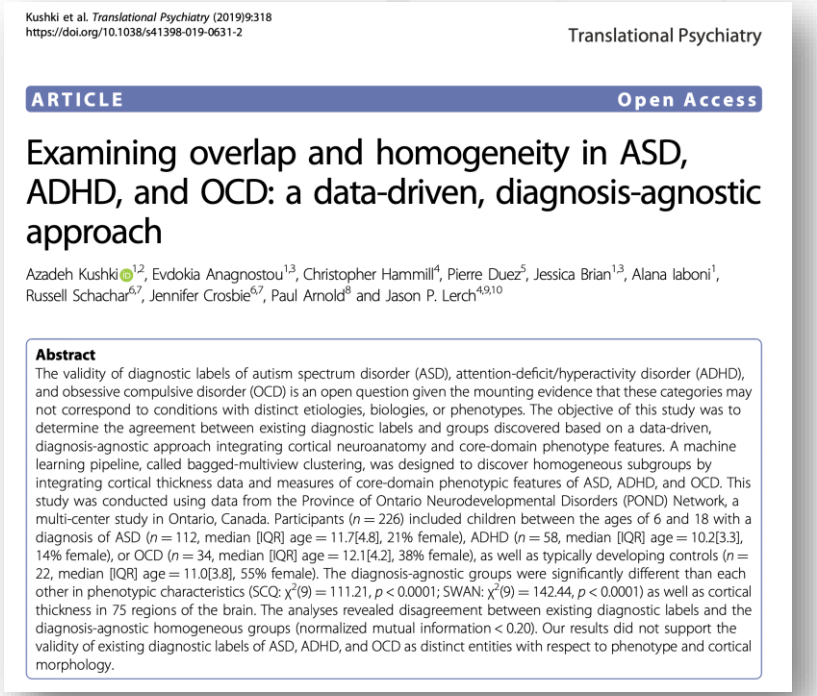
Approach

- Multi-modal data to see which combination of modalities would best predict response to escitalopram after two weeks
- 3 modalities: clinical, imaging, molecular
- 192 participants
- 5 ML methods tested: Elastic net, GBM, Random forest, SVM, Naïve Bayes

Results

- Clinical and molecular performed best (73% AUC)
- Using Elastic net
- Combining early and late treatment data worked best

Data driven diagnostic grouping in Neurodevelopment



Kushki et al. *Translational Psychiatry* (2019)9:318
https://doi.org/10.1038/s41398-019-0631-2

Translational Psychiatry

ARTICLE Open Access

Examining overlap and homogeneity in ASD, ADHD, and OCD: a data-driven, diagnosis-agnostic approach

Azadeh Kushki[1,2], Evdokia Anagnostou[1,3], Christopher Hammill[4], Pierre Duez[5], Jessica Brian[1,3], Alana Iaboni[1], Russell Schachar[6,7], Jennifer Crosbie[6,7], Paul Arnold[8] and Jason P. Lerch[4,9,10]

Abstract

The validity of diagnostic labels of autism spectrum disorder (ASD), attention-deficit/hyperactivity disorder (ADHD), and obsessive compulsive disorder (OCD) is an open question given the mounting evidence that these categories may not correspond to conditions with distinct etiologies, biologies, or phenotypes. The objective of this study was to determine the agreement between existing diagnostic labels and groups discovered based on a data-driven, diagnosis-agnostic approach integrating cortical neuroanatomy and core-domain phenotype features. A machine learning pipeline, called bagged-multiview clustering, was designed to discover homogeneous subgroups by integrating cortical thickness data and measures of core-domain phenotypic features of ASD, ADHD, and OCD. This study was conducted using data from the Province of Ontario Neurodevelopmental Disorders (POND) Network, a multi-center study in Ontario, Canada. Participants ($n = 226$) included children between the ages of 6 and 18 with a diagnosis of ASD ($n = 112$, median [IQR] age = 11.7[4.8], 21% female), ADHD ($n = 58$, median [IQR] age = 10.2[3.3], 14% female), or OCD ($n = 34$, median [IQR] age = 12.1[4.2], 38% female), as well as typically developing controls ($n = 22$, median [IQR] age = 11.0[3.8], 55% female). The diagnosis-agnostic groups were significantly different than each other in phenotypic characteristics (SCQ: $\chi^2(9) = 111.21$, $p < 0.0001$; SWAN: $\chi^2(9) = 142.44$, $p < 0.0001$) as well as cortical thickness in 75 regions of the brain. The analyses revealed disagreement between existing diagnostic labels and the diagnosis-agnostic homogeneous groups (normalized mutual information < 0.20). Our results did not support the validity of existing diagnostic labels of ASD, ADHD, and OCD as distinct entities with respect to phenotype and cortical morphology.

POND

Approach

- Cross clinical diagnostic labelled population was combined from ASD, OCD, and ADHD
- 2 modalities: cortical thickness (imaging), phenotypic
- 226 participants data
- Bagged multi-view clustering

Results

- Clusters were significantly different from each other (SWAN 142.22 p<0.0001)
- Clusters did not match traditional clinical labels

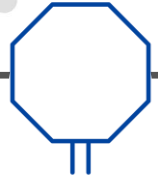
Health Data | Challenges with the Centralized Approach

DOESN'T GENERALIZE

- Models often over fit
- Models can drift as new cases arise
- Models can drift due to practice changes

UNDERREPRESENTED POPULATIONS

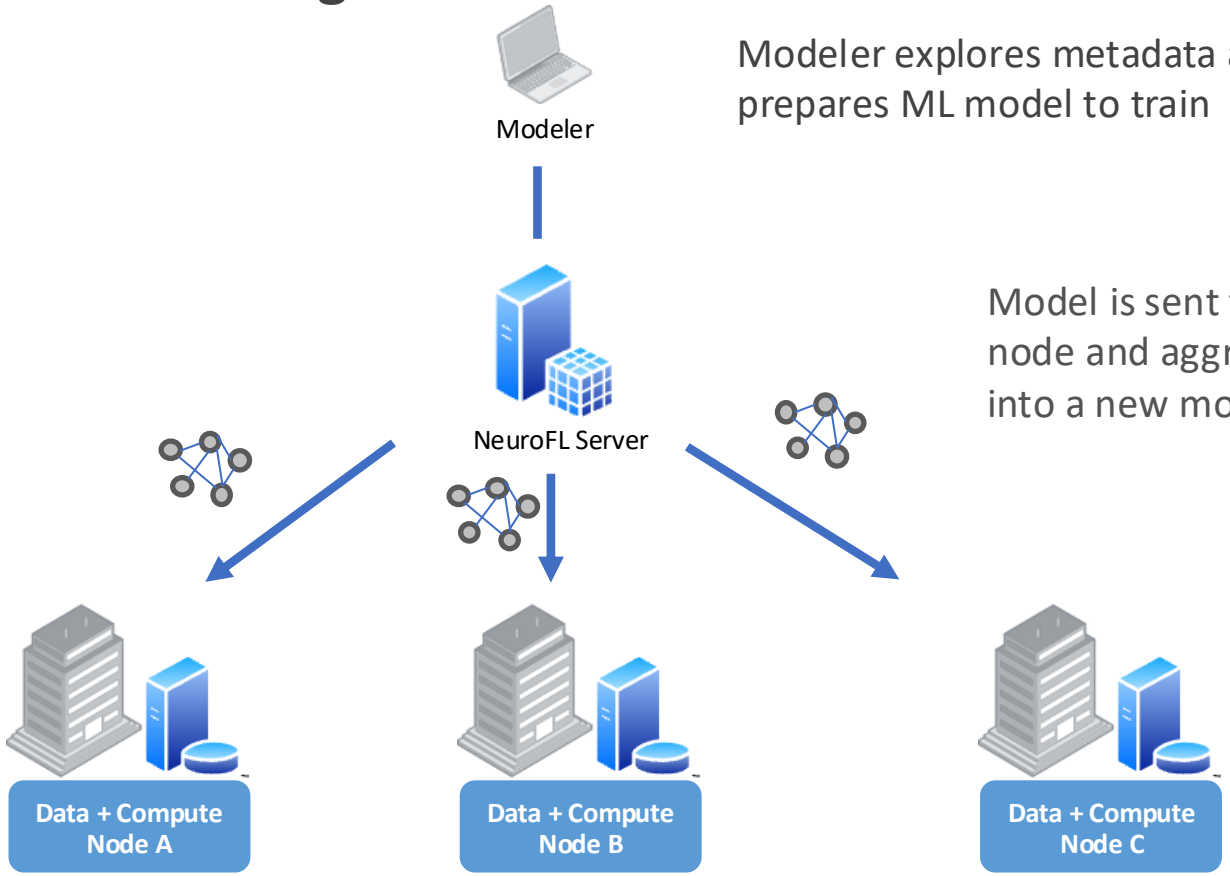
- Data doesn't represent patient populations
- Subgroups of patients cannot be explained
- Outcomes can be biased



CAN'T ALWAYS MOVE

- Data sharing agreements take time
- Data may be too large & costly to move
- Data may be too sensitive to share

Federated Learning?



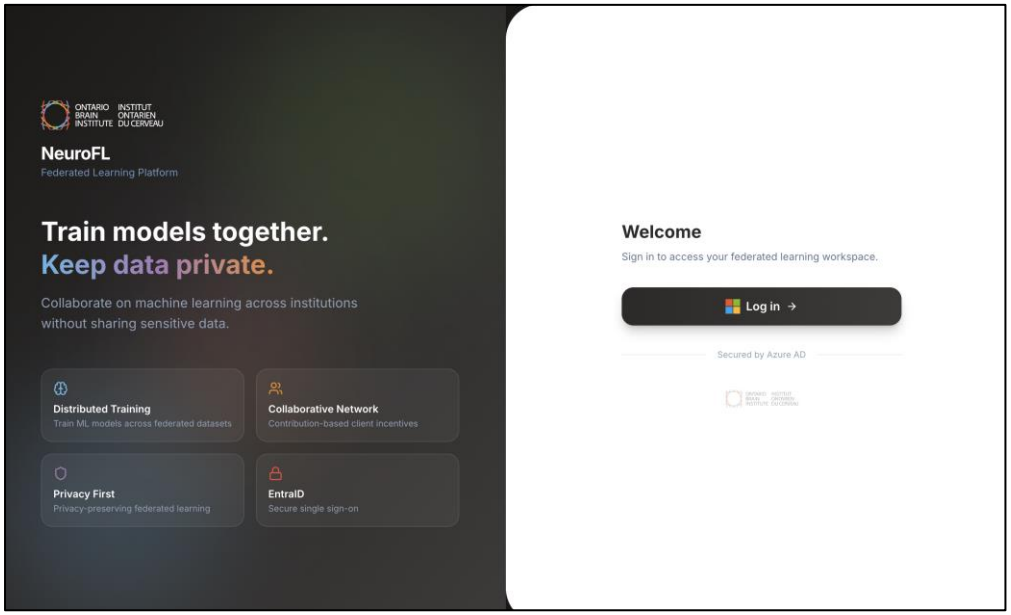
Modeler explores metadata and prepares ML model to train

Model is sent to each NeuroFL node and aggregates results into a new model

Data always stay onsite

NeuroFL

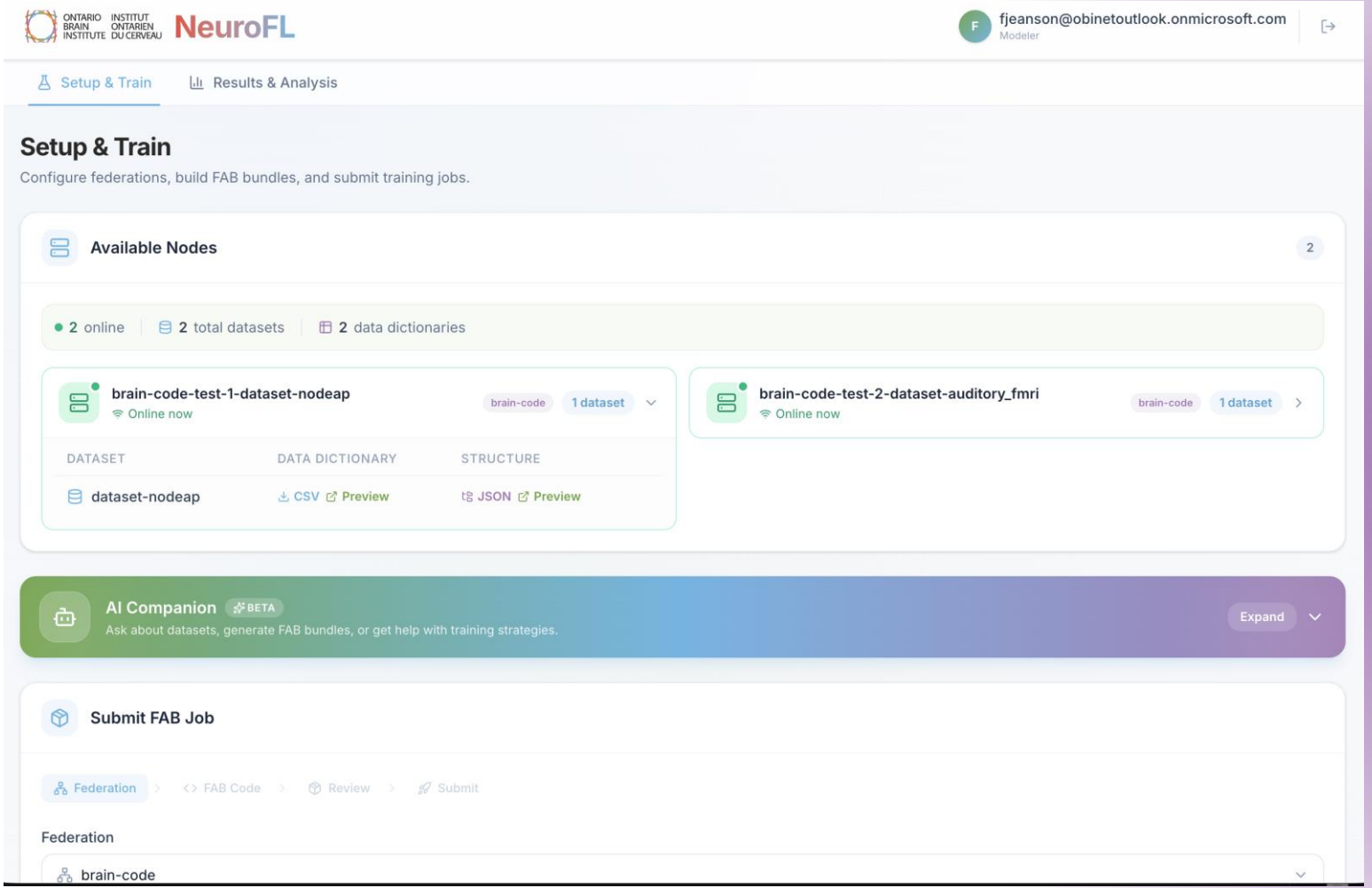
A Federated Data Network for Brain Health



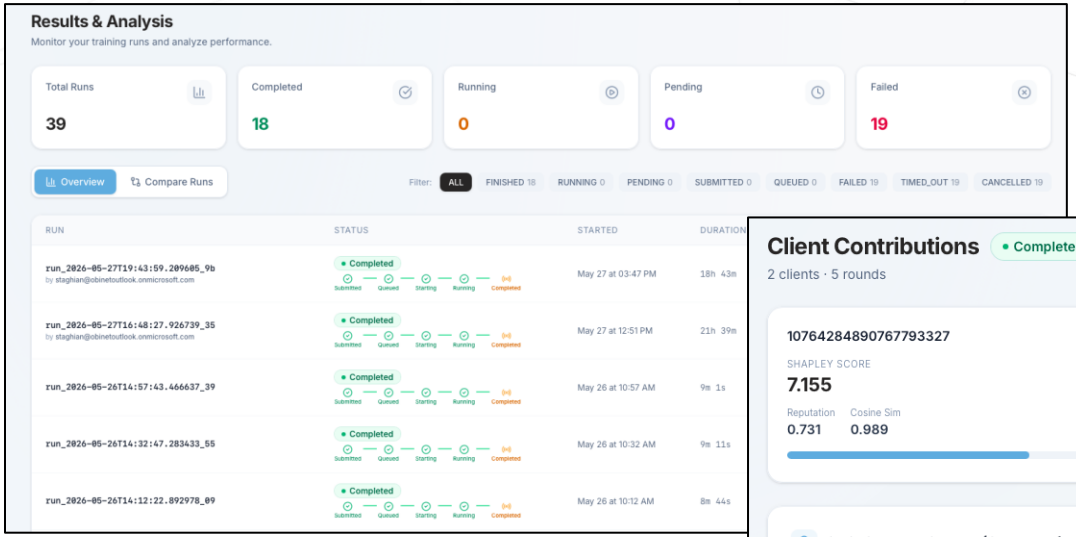
NeuroFL

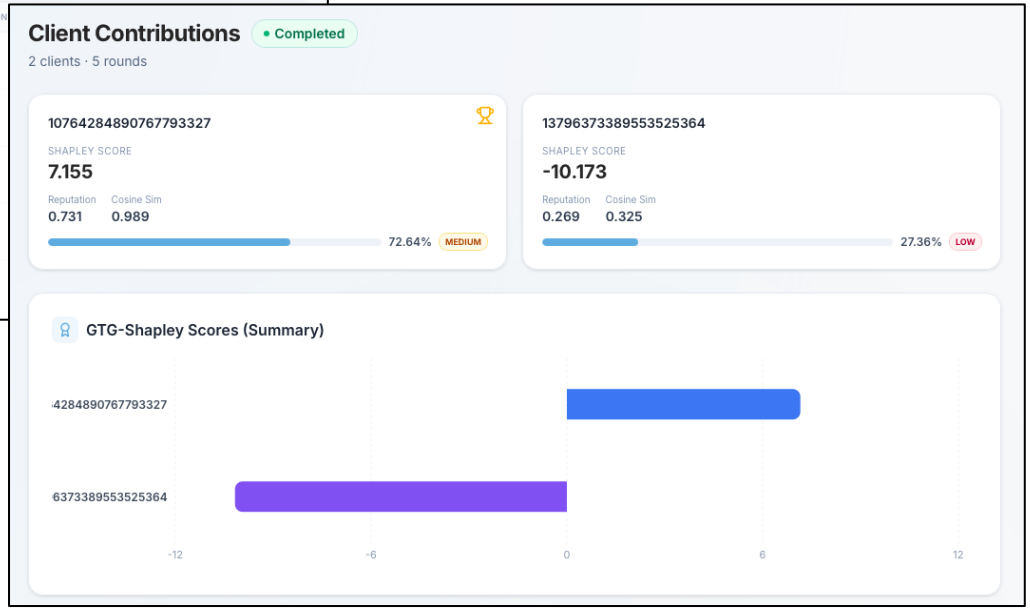
- Secure federated learning workflows
- Protects data sovereignty and privacy of data for greater protections
- Train models on more data
- Broader population samples for greater AI generalizability and representation
- Multi-site collaboration with real-time training monitoring

Ontario Brain Institute
Centre for Analytics


ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU
NeuroFL
fjeanson@obinetoutlook.onmicrosoft.com
Modeler
Setup & Train
Results & Analysis
Setup & Train
Configure federations, build FAB bundles, and submit training jobs.
Available Nodes
2
2 online
2 total datasets
2 data dictionaries
brain-code-test-1-dataset-nodeap
Online now
brain-code
1 dataset
DATASET
DATA DICTIONARY
STRUCTURE
dataset-nodeap
CSV Preview
JSON Preview
brain-code-test-2-dataset-auditory_fmri
Online now
brain-code
1 dataset
AI Companion
BETA
Ask about datasets, generate FAB bundles, or get help with training strategies.
Expand
Submit FAB Job
Federation
FAB Code
Review
Submit
Federation
brain-code

NeuroFL Results & Node Contributions

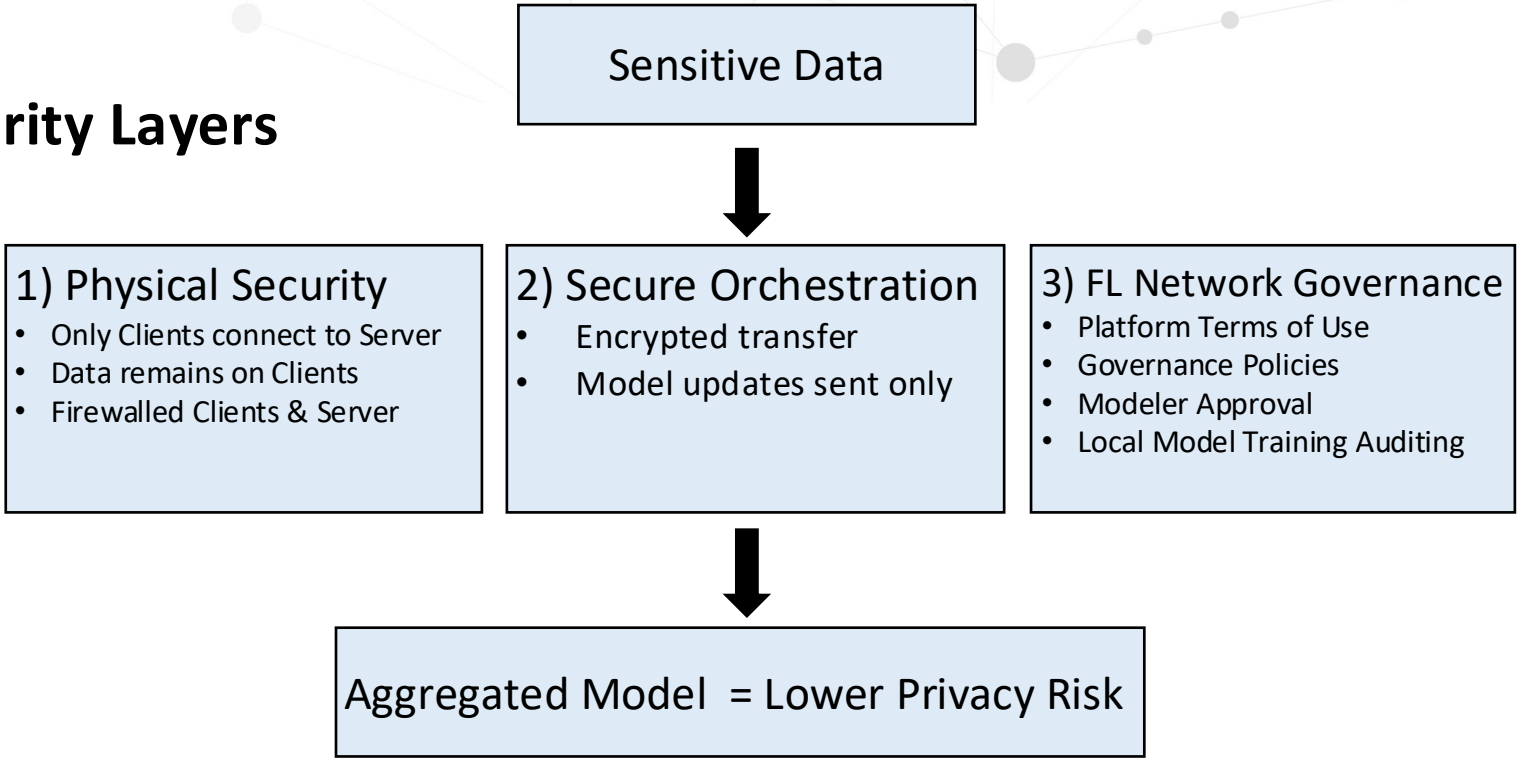




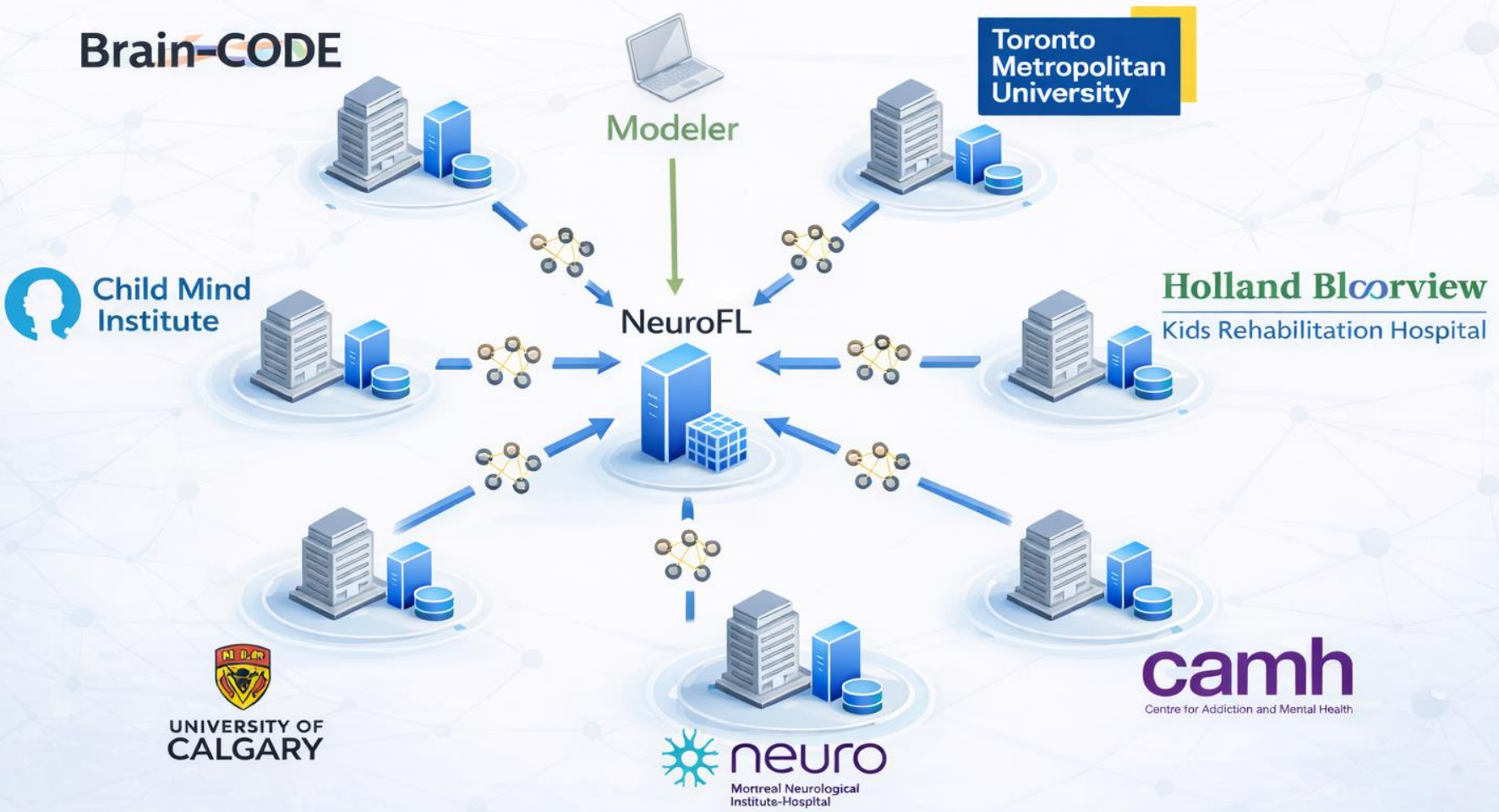
Ontario Brain Institute
Centre for Analytics

NeuroFL Security & Privacy

Security Layers



NeuroFL | A data intelligence network for Brain Health



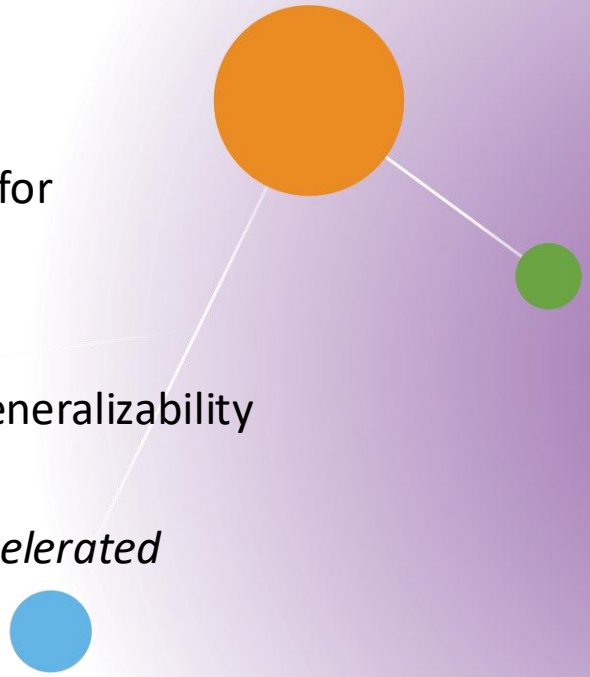
NeuroFL Technical Collaborators

- Patrick Foley, Flower Labs
  - Flower Integration and Features
- JB Poline & team, Origami Lab, McGill
  - Nipoppy/Neurobagel pipelines
- Nils Forkert, MIP Lab
  - FL, FM, Synth data
- David Emerson, Vector Institute
  - FL4Health team
- Spyros Bakas, Bakas Lab
  - FL, LLM

NeuroFL

Benefits

- *Data don't move…*
- Preserve the sovereignty and privacy of data for greater protections
- Train models on larger amounts of data
- Broaden population samples for greater AI generalizability and representation
- *An ecosystem of connected data nodes to accelerated brain health discovery*



Thank you!

Francis Jeanson, PhD
fjeanson@braininstitute.ca

ONTARIO BRAIN INSTITUTE | INSTITUT ONTARIEN DU CERVEAU